



## City Research Online

### City, University of London Institutional Repository

---

**Citation:** Miranda, M. D. M., Nielsen, J. P. and Sperlich, S. (2009). One Sided Crossvalidation for Density Estimation. In: Gregoriou, G.N. (Ed.), Operational Risk Towards Basel III: Best Practices and Issues in Modeling, Management and Regulation. (pp. 177-196). New Jersey: John Wiley and Sons.

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

---

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/4635/>

**Link to published version:**

**Copyright:** City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

**Reuse:** Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

---

---



# One Sided Cross Validation for Density Estimation with an Application to Operational Risk

**María Dolores Martínez Miranda**

Universidad de Granada

Departamento de Estadística e I.O.

Campus Fuentenueva, E - 18071 Granada

**Jens Perch Nielsen**

University of Copenhagen and Festina Lente,

Sondersovej 52

DK - 2820 Copenhagen, Denmark

**Stefan Sperlich**

Georg-August Universität Göttingen

Institut für Statistik und Ökonometrie

Platz der Göttinger Sieben 5, D - 37073 Göttingen

## Abstract

We introduce one-sided cross-validation to nonparametric kernel density estimation. The method is more stable than classical cross-validation and it has a better overall performance comparable to what we see in plug-in methods. One-sided cross-validation is a more direct data driven method than plug-in methods with weaker assumptions of smoothness since it does not require a smooth pilot with consistent second derivatives. Our conclusions for one-sided kernel density cross-validation are similar to the conclusions obtained by Hart and Li (1998) when they introduced one-sided cross-validation in the regression context. An extensive simulation study confirms that our one-sided cross-validation clearly outperforms the simple cross validation. We conclude with real data applications.<sup>1</sup>

**Keywords:** bandwidth choice, cross-validation, plug-in, nonparametric estimation

---

<sup>1</sup>This research was financially supported by knowledge company Festina Lente and the Dirección General de Investigación del Ministerio de Ciencia y Tecnología under research grant SEJ2004-04583/ECON.

# 1 Introduction

Suppose we have observed data  $X_1, X_2, \dots, X_n$  that are assumed independent identically distributed with common density function,  $f(\cdot)$ . We want to estimate this common density nonparametrically using the standard kernel estimator:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad (1)$$

where  $K$  is the kernel function and  $h$  is the bandwidth parameter. Our problem is to find a reliable data driven estimator of the bandwidth. We would like to use the popular and widely used least squares cross-validation proposed of Rudemo (1982) and Bowman (1984). We do, however, worry about the well known lack of stability of this method, see e.g. Wand and Jones (1995). Many alternatives have been proposed to the intuitively appealing method of cross-validation. For example the wide range of so called plug-in methods aiming at estimating the minimizer of the integrated squared error. However, all these plug-in methods require a pilot estimator to be plugged in. We prefer a direct and immediate method like cross-validation without extra complications with pilot estimators and without the extra smoothing assumptions required to assure that the pilot estimator works well. In regression an appealing improvement of standard cross-validation exists, namely the so called one-sided cross-validation that simply is the cross-validation procedure based on the one-sided kernel version of the original kernel  $K(\cdot)$ . The estimated bandwidth coming from this procedure is then readjusted by a simple constant only depending on the kernel, see Hart and Yi (1998) and Yi (1996, 2001, 2004). The surprising fact is that this one-sided procedure is much more stable than the original cross-validation procedure and that it in many ways behaves similar to the plug-in method without having its vices: the complicated pilot estimator and the added smoothness assumptions. In this paper we introduce one-sided cross-validation for kernel density estimation and we show through simulations that one-sided cross-validation is performing better, is more stable and have a similar performance to plug-in, also in the kernel density case. Its performance and superiority will be shown via simulation studies and real data applications.

## 2 The one-sided cross-validation method for density estimation

One commonly used measure of the performance of  $\hat{f}_h$  is the Mean Integrated Squared Error (MISE), defined by

$$\text{MISE}(h) = \int \mathbb{E}[\{\hat{f}_h(x) - f(x)\}^2] dx.$$

Let's denote by  $h_0$  the minimizer of  $\text{MISE}(\cdot)$ . This is the optimal bandwidth that plug-in methods aim at estimating. Another performance measure to consider is the data dependent Integrated Squared Error, defined by

$$\text{ISE}(h) = \int \left( \hat{f}_h(x) - f(x) \right)^2 dx$$

with the optimal (random) bandwidth,  $\hat{h}_0$  as minimizer.

This is the optimal bandwidth that cross-validation aims at estimating. However, theoretical studies have shown that standard cross-validation is so unstable that plug-in methods do better at estimating  $\hat{h}_0$  than cross-validation does, even though plug-in methods really aim at estimating  $h_0$ .

### 2.1 Ordinary least squares cross-validation

Cross-validation is probably still the most popular automatic bandwidth selection method. Its intuitive definition and its practical data driven flavor makes up for its lack of stability in the eyes of many practitioners. Also cross-validation immediately generalizes to most statistical smoothing problems. Plug-in methods are only well defined for a narrow range of statistical problems and even here debate over which pilot estimator to use makes practitioners turn to cross-validation.

Least squares cross-validation was proposed by Rudemo (1982) and Bowman (1984), who estimated  $\hat{h}_0$  by minimizing the criterion,

$$\text{CV}(h) = \int \hat{f}_h^2(x) dx - 2n^{-1} \sum_{i=1}^n \hat{f}_{h,-i}(X_i), \quad (2)$$

where  $\hat{f}_{h,-i}$  is the density estimator obtained by leaving out the observation  $X_i$ . Let  $\hat{\hat{h}}$  be this classical cross-validation bandwidth estimator.

Hall (1983) showed that the cross-validation bandwidth is a consistent estimate of the optimal bandwidth  $\hat{h}_0$ , and its asymptotic normality was established in Hall and Marron (1987). They pointed out the lack of stability of classical cross-validation. Härdle, Hall and Marron (1988) showed the equivalent result for the regression context. The cross-validation bandwidth also tends to be undersmoothing in many practical applications. There has therefore been a number of studies on more stable bandwidth selectors. Most of them related to the plug-in method. For example the plug-in method of Sheather and Jones (1986), biased cross-validation by Scott and Terrell (1987) smoothed cross-validation by Hall, Marron and Park (1989) and the stabilized bandwidth selector rule by Chiu (1991).

## 2.2 One-sided cross-validation

Hart and Yi (1998) used local linear regression when introducing one-sided cross-validation in the regression context. They did this for a good reason since a good boundary correction method is crucial for the one-sided procedure. We therefore combine our one-sided cross-validation method with the local linear density estimator of Jones (1993). In density estimation the local linear density estimator is identical to the standard kernel density estimator away from boundaries, see below.

Let  $K(\cdot)$  be any common symmetric kernel function and let consider its (left)one-sided version,

$$\bar{K}(u) = \begin{cases} 2K(u) & \text{if } u < 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

Consider the one-sided density estimator,  $\hat{f}_{left,b}$  based on the one-sided kernel  $\bar{K}$  and bandwidth  $b$ . we define the one-sided versions of the error measures ISE and MISE calling them OISE and MOISE. Define also  $\hat{b}_0$  and  $b_0$ , their minimizers (respectively).

We also have the following assumptions on the kernel:  $\mu_0(K) = 1$ ,  $\mu_1(K) = 0$  and  $\mu_2(K) < \infty$ , where  $\mu_l(K) = \int u^l K(u) du$  ( $l = 0, 1, 2$ ).

The one-sided cross-validation criterion like is defined as

$$\text{OSCV}(b) = \int \hat{f}_{left,b}^2(x) dx - 2n^{-1} \sum_{i=1}^n \hat{f}_{left,b}(X_i), \quad (4)$$

with  $\hat{b}$  as its minimizer.

With these definitions, the one-sided cross-validation bandwidth is based on  $\widehat{\widehat{b}}$  but it has to be adjusted by a constant to become an estimator of the bandwidth for the original kernel density estimator. Let us define

$$\widehat{h}_{OSCV} := C\widehat{\widehat{b}}, \quad (5)$$

where the constant,  $C$ , will be the ratio of the optimal bandwidth (in MISE sense) of the density estimator ( $\widehat{f}_h$ ) to the optimal bandwidth (in MOISE sense) of the one-sided density estimator ( $\widehat{f}_{left,b}$ ), i.e.

$$C = \frac{h_0}{b_0}.$$

Note that this constant does not depend on the underlying density. To be able to do the one-sided kernel estimator we consider the local linear boundary corrected density estimator of Jones (1993) and Cheng (1997a,1997b). Consider the minimization problem:

$$\min_{\beta_0, \beta_1} \left[ \int \{f_n(u) - \beta_0 - \beta_1(u - x)\}^2 K\left(\frac{u - x}{h}\right) du \right],$$

where  $f_n(u) = n^{-1} \sum_{i=1}^n \mathbf{1}_{\{u=X_i\}}$  is the empirical density function. Then, the local linear estimator is defined by the equivalent-kernel expression

$$\widehat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K^*\left(\frac{X_i - x}{h}\right) \quad (6)$$

with the equivalent kernel

$$\begin{aligned} K^*(u) &= e_1^T S^{-1}(1, u)^T K(u) = \\ &= \frac{\mu_2(K) - \mu_1(K)u}{\mu_0(K)\mu_2(K) - (\mu_1(K))^2} K(u), \end{aligned} \quad (7)$$

being  $e_1 = (1, 0)^T$  and  $S = (\mu_{i+j-2})_{0 \leq i, j \leq 2}$ .

We define the operator  $R(g) = \int (g(x))^2 dx$ , for a generic squared integrable function  $g$ . Then the optimal bandwidths for local linear estimator is given by

$$h_0 = \left( \frac{R(K)}{(\mu_2(K))^2 R(f'')} \right)^{1/5} n^{-1/5}, \quad (8)$$

for the ordinary local linear estimator and

$$b_0 = \left( \frac{R(\bar{K}^*)}{(\mu_2(\bar{K}^*))^2 R(f'')} \right)^{1/5} n^{-1/5}, \quad (9)$$



for the one-sided version, where  $\bar{K}^*$  is the one-sided version of the equivalent kernel (7), i.e.,

$$\bar{K}^*(u) = \frac{\mu_2(K) - u \left( 2 \int_{-\infty}^0 t K(t) dt \right)}{\mu_2(K) - \left( 2 \int_{-\infty}^0 t K(t) dt \right)^2} 2K(u) \mathbf{1}_{\{u < 0\}}. \quad (10)$$

The adjusting constant becomes then

$$C = \left( \frac{R(K)}{R(\bar{K}^*)} \frac{\mu_2(\bar{K}^*)^2}{\mu_2(K)^2} \right)^{1/5}, \quad (11)$$

which obviously is a feasible number.

### 3 Asymptotic theory

The theoretical justification for the stability of one-sided cross-validation seems to come from the fact that the variation of one-sided cross-validation around that optimal bandwidth it is aiming at estimating is much smaller than the variation of ordinary cross-validation around its optimal bandwidth. We will carry out the details of this argument below following Hall and Marron (1987).

#### Assumptions

- (A1) The density,  $f$ , is bounded and twice differentiable,  $f'$  and  $f''$  are bounded and integrable, and  $f''$  is uniformly continuous.
- (A2) The kernel  $K$  is compactly supported, symmetric density function on  $\mathbb{R}$  with Hölder-continuous derivative,  $K'$ , and satisfies that  $\mu_2(K) \neq 0$ .

Consider the following additional definitions and notation, assumed that (A1)-(A2) hold:

Let define  $W(u) = -uK'(u)$  and the one-sided version,  $\bar{W}^*(u) = -u\bar{K}^{*'}(u)\mathbf{1}_{\{u < 0\}}$ , under assumption (A2) these functions are kernels which integrate one and verify that  $\mu_1(W) = \mu_1(\bar{W}^*) = 0$ .

Let be the constants:

$$c_0 = (R(K)/(\mu_2(K)^2 R(f''^{1/5}))$$

and

$$c_1 = 2c_0^{-3} R(K) + 3(\mu_2(K))^2 R(f'') c_0^2.$$

And the one-sided versions:  $\bar{c}_0 = (R(\bar{K}^*)/(\mu_2(\bar{K}^*)^2 R(f'^{1/5}))$  and  $\bar{c}_1 = 2\bar{c}_0^3 R(\bar{K}^*) + 3(\mu_2(\bar{K}^*))^2 R(f'')\bar{c}_0^2$ .

Let us define the variance terms:

$$\sigma_c^2 = (2/c_0)^3 R(f)R(W) + (2\mu_2(K)c_0)^2 \left\{ \int (f''^2 f - \left( \int f'' f \right)^2) \right\} \quad (12)$$

and

$$\bar{\sigma}_{oc}^2 = (2/\bar{c}_0)^3 R(f)R(\bar{W}^*) + (2\mu_2(\bar{K}^*)\bar{c}_0)^2 \left\{ \int (f''^2 f - \left( \int f'' f \right)^2) \right\}. \quad (13)$$

Observe that the difference  $\int (f''^2 f - \left( \int f'' f \right)^2)$  is the variance of  $f''(X)$ . It will be denoted by  $V(f'')$  in the following.

Under conditions (A1) and (A2) Hall and Marron (1987) demonstrated that

$$n^{3/10}(\hat{h} - \hat{h}_0) \longrightarrow \mathcal{N}(0, \sigma_c^2 c_1^{-2}). \quad (14)$$

An immediate application of Hall and Marron (1987) gives the following result allowing us to compare the variation of one-sided cross-validation to the variation of standard cross-validation:

**Theorem 1.** *Under conditions (A1) and (A2),*

$$n^{3/10}(\hat{h}_{OSCV} - \hat{C}b_0) \longrightarrow \mathcal{N}(0, C^2 \bar{\sigma}_{oc}^2 \bar{c}_1^{-2}). \quad (15)$$

The gain in reduction of the variation can be approximated as follows.

**Remark 1.** *The ratio of the asymptotic variance of one-sided cross-validation to standard cross-validation is given by the ratio of the asymptotic variance from (14) and the asymptotic variance of (1):*

$$G_{oc} = C^2 \left( \frac{c_1}{\bar{c}_1} \right)^2 \frac{\bar{\sigma}_{oc}^2}{\sigma_c^2}. \quad (16)$$

*The reduction of variance for the Epanechnikov kernel and Gaussian kernel are given by:*

$$G_{oc}^{Ep} = \frac{0.530605928 R(f)R(f'') + 0.117383673 V(f'')}{1.418004931 R(f)R(f'') + 0.47266831 V(f'')}. \quad (17)$$

and

$$G_{oc}^{Ga} = \frac{0.6913873 R(f)R(f'') + 0.6173955 V(f'')}{17.094364 R(f)R(f'') + 1.363272 V(f'')}. \quad (18)$$

*The variance reduction is at least 35% for the Epanechnikov kernel and at least 50% for the gaussian kernel.*

## 4 Finite Sample Performance

The small sample performance of one-sided kernel density estimation is compared to its most immediate competitors, classical cross-validation and plug-in. The here chosen plug-in method is also called "refined plug-in" or often referred to as Sheather-Jones bandwidth; details are given below. The performance is compared by integrated squared error (ISE), where we consider two measures. The classical where the ISE is calculated for each sample and averaged ( $m_3$ ) and a new - and perhaps better measure - where we the integrated squared error of ISE's ( $m_1$ ) and the  $L_1$ -distance ( $m_2$ ) are calculated. The latter measures therefore take variability of the ISE's into account and penalize bandwidth selectors that often do well but once in a while fail completely. We also calculate the bias of the bandwidth selectors ( $m_5$ ) and the volatility of the ISE's ( $m_4$ ). These numbers will help us explaining why one-sided cross-validation does better than classical cross-validation. Concretely, given a bandwidth estimate,  $\hat{h}$ , the considered criteria are the following:

$$m_1 = \text{mean}(\{\text{ISE}(\hat{h}) - \text{ISE}(\hat{h}_0)\}^2),$$

$$m_2 = \text{mean}(|\text{ISE}(\hat{h}) - \text{ISE}(\hat{h}_0)|),$$

$$m_3 = \text{mean}(\text{ISE}(\hat{h})),$$

$$m_4 = \text{std}(\text{ISE}(\hat{h})),$$

and

$$m_5 = \text{mean}(\hat{h} - \hat{h}_0).$$

For brevity we will concentrate on kernel density estimation with the local linear Epanechnikov kernel.

The plug-in bandwidth  $h_0$  is calculated from equation (8). Here,  $R(K)$  and  $\mu_2(K)$  are known whereas  $R(f'')$  has to be estimated with a prior bandwidth  $g_p$ . To this aim, take Silverman's rule of thumb bandwidth  $g_p$  for Gaussian kernels where the standard deviation of  $X$  is estimated by the minimum of two methods: the moment estimate  $s_n$  and the interquartile range  $IR_X$  divided by 1.34, i.e.  $g_S = 1.06 \min\{IR_X 1.34^{-1}, s_n\} n^{-1/5}$ . Then, as the Quartic kernel  $K_Q$  comes close to the Epanechnikov but allows for estimating the second derivative, we normalize  $g_S$  by the factors of the canonical kernel (Gaussian to Quartic) and adjust for the slower rate ( $n^{-1/9}$ ) needed to estimate second derivatives, i.e.

$$g_p = g_S \frac{2.0362}{0.7764} n^{1/5-1/9}.$$

Next,

$$\widehat{R}(f'') = R(\hat{f}'') - \frac{1}{ng_p^5} R(K_Q'')$$

to correct for the bias inherited by

$$\hat{f}''(x) = \frac{1}{ng_p^3} \sum_{i=1}^n K_Q'' \left( \frac{x - X_i}{g_p} \right) .$$

In simulation studies not shown here this prior choice turned out to perform better than any modification we have tried, at least for the estimation problems discussed in the following.

## 4.1 Data Generating Processes and Numerical Results

As data generating process (DGP) we have considered a huge bunch of normal, gamma and mixed densities from which we will concentrate on the following six:

1. a simple normal distribution,  $N(0.5, 0.2^2)$ ,
2. a mixture of two normals which were  $N(0.35, 0.1^2)$  and  $N(0.65, 0.1^2)$ ,
3. a gamma,  $Gamma(a, b)$  with  $b = 1.5$ ,  $a = b^2$  applied on  $5x$  with  $x \in \mathbb{R}$ ,  
i.e.

$$f(x) = 5 \frac{b^a}{\Gamma(a)} (5x)^{a-1} e^{-5xb},$$

4. a mixture of two gamma,  $Gamma(a_j, b_j)$ ,  $a_j = b_j^2$ ,  $b_1 = 1.5$ ,  $b_2 = 3$   
applied on  $6x$ , i.e.

$$f(x) = \frac{6}{2} \sum_{j=1}^2 \frac{b_j^{a_j}}{\Gamma(a_j)} (6x)^{a_j-1} e^{-6xb_j}$$

giving one mode and a plateau,

5. a mixture of three gamma,  $Gamma(a_j, b_j)$ ,  $a_j = b_j^2$ ,  $b_1 = 1.5$ ,  $b_2 = 3$ ,  
and  $b_3 = 6$  applied on  $8x$  giving two bumps and one plateau,
6. and a mixture of three normals, namely  $N(0.25, 0.075^2)$ ,  $N(0.5, 0.075^2)$   
and  $N(0.75, 0.075^2)$  giving three clear modes.

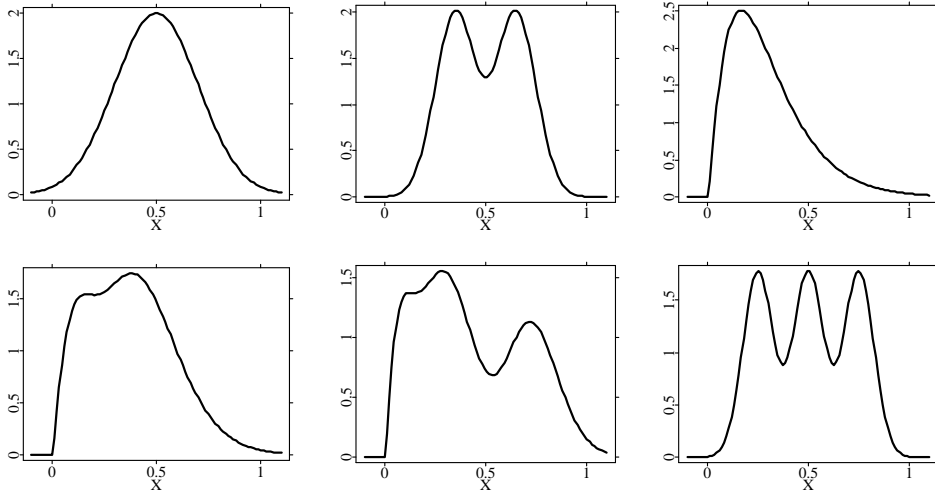


Figure 1: The six data generating densities: design 1 to 6 from the upper left to the lower right.

As you can see in Figure 1 all six models have the main mass in  $[0, 1]$ . You can see also that we have mostly densities with exponentially decreasing tails so that in this simulation study we will disregard the possible use of boundary correcting kernels. Moreover, we assume that the empirical researcher has no knowledge on possible boundaries. For the six models we have considered sample sizes:  $n = 50, 100$  and  $200$ , and 250 repetitions (simulation runs) for each model and each sample size.

The results of the six densities are collected in Table 1 to Table 3. We see that one-sided cross-validation does better and is much more stable than classical cross-validation on most of our performance measures (see especially  $m_1$  and  $m_2$ ). Therefore, the relative improvement in performance is even bigger with our main performance measure  $m_1$ , compare Remark 1. One can see that the price for stability (compare e.g.  $m_4$ ) of both the one-sided cross-validation and the plug-in method is a tendency to overestimate the bandwidth a little bit, see  $m_5$ . However, the stability more than makes up for this bias and the overall performance of both these methods tend to be better than the performance of classical cross-validation. To see this, recall that the measures of interest for the practitioner are  $m_1$  to  $m_3$ . The conclusion is that one-sided cross-validation performs similar to - sometimes worse, sometimes better - as the plug-in method. Our results therefore parallel the results of Hart and Yi (1998) in the regression context.

$n$	Criteria	Design 1			Design 2		
		CV	OSCV	Plug-in	CV	OSCV	Plug-in
50	$m_1$	,0112	,0004	3e-05	,0171	,0012	,0016
	$m_2$	,0407	,0064	,0029	,0552	,0263	,0357
	$m_3$	,0781	,0438	,0403	,1112	,0823	,0917
	$m_4$	,1028	,0372	,0305	,1214	,0351	,0197
	$m_5$	-,0225	,0136	,0193	,0071	,0500	,0803
100	$m_1$	,0028	8e-05	1e-05	,0019	,0004	,0009
	$m_2$	,0234	,0039	,0014	,0246	,0142	,0288
	$m_3$	,0479	,0285	,0260	,0590	,0486	,0633
	$m_4$	,0530	,0209	,0185	,0450	,0241	,0137
	$m_5$	-,0234	,0088	,0160	,0067	,0308	,0763
200	$m_1$	,0017	3e-05	1e-06	,0021	,0007	,0005
	$m_2$	,0110	,0024	,0007	,0159	,0084	,0213
	$m_3$	,0255	,0170	,0152	,0384	,0309	,0438
	$m_4$	,0417	,0125	,0108	,0448	,0287	,0116
	$m_5$	-,0136	,0034	,0127	,0003	,0145	,0671

Table 1: Criteria values for designs 1 and 2

## 5 Practical Remarks and Data Applications

### 5.1 Data Transformation and Boundary Correction

In our application we estimate data belonging to the interval (0,1) because we want to apply the transformation approach of Buch-Larsen et al. (2005) to estimate some loss distributions of operational risk. While the transformation methodology of Buch-Larsen et al. (2005) have proved to be extremely efficient and beat its direct competitors in the extensive simulation study of that paper, the bandwidth selection part of that paper is not really very sophisticated. It is just the simplest possible bandwidth selector: Silverman's rule of thumb. Recall that if the prior information of facing a one mode distribution is available, Silverman's rule of thumb may give nice plots but is generally much too coarse for a detailed data analysis. So, while the transformation method of Buch-Larsen et al. (2005) already has shown its usefulness it clearly needs to be updated by a better bandwidth selection method. We use cross-validation, refined plug-in, Silverman's rule of thumb and our new

		Design 3			Design 4		
$n$	Criteria	CV	OSCV	Plug-in	CV	OSCV	Plug-in
50	$m_1$	,0164	,0026	,0019	,0062	,0009	6e-05
	$m_2$	,0564	,0339	,0342	,0343	,0096	,0053
	$m_3$	,1316	,1092	,1095	,0823	,0576	,0534
	$m_4$	,1206	,0557	,0457	,0756	,0353	,0215
	$m_5$	-5e-05	,0456	,0669	-,0081	,0335	,0398
100	$m_1$	,0018	,0006	,0010	,0007	8e-05	4e-05
	$m_2$	,0227	,0196	,0269	,0126	,0056	,0047
	$m_3$	,0698	,0667	,0741	,0466	,0397	,0387
	$m_4$	,0480	,0329	,0306	,0276	,0158	,0134
	$m_5$	-,0008	,0388	,0666	-,0035	,0327	,0395
200	$m_1$	,0030	,0008	,0005	,0022	,0008	3e-05
	$m_2$	,0168	,0126	,0193	,0130	,0061	,0040
	$m_3$	,0495	,0453	,0520	,0365	,0296	,0275
	$m_4$	,0559	,0318	,0200	,0460	,0292	,0070
	$m_5$	-,0038	,0245	,0593	-,0056	,0302	,0393

Table 2: Criteria values for designs 3 and 4

one-sided cross validation estimator as our selection rule. We conclude that while the other estimators grossly oversmooth or undersmooth what seems to be appropriate, the one-sided cross validation seems to work very well in practice.

First a few words on the actual transformation. In our two considered cases the modified Champerknowne distribution of Buch-Larsen et al. (2005) actually simplifies since their parameter  $c$  is estimated to zero. This implies that the transformation is a special case of the original Champerknown distribution - Champerknown (1936,1952) and from a transformation point of view identical to the Möbius transformation used for the exact same purpose by Clements, Hurd and Lindsay (2003). The simple Champerknown distribution has cumulated distribution function

$$T(x) = x^\alpha(x^\alpha + M^\alpha)^{-1}$$

with density

$$t(x) = \alpha x^{\alpha-1} M^\alpha (x^\alpha + M^\alpha)^{-2} ,$$

where  $M$  and  $\alpha$  will be estimated via maximum likelihood on the original data.

		Design 5			Design 6		
$n$	Criteria	CV	OSCV	Plug-in	CV	OSCV	Plug-in
50	$m_1$	,0034	,0042	,0001	,0095	,0055	,0055
	$m_2$	,0283	,0239	,0088	,0455	,0659	,0669
	$m_3$	,0810	,0765	,0615	,1173	,1377	,1387
	$m_4$	,0591	,0664	,0222	,0896	,0159	,0126
	$m_5$	,0023	,0860	,0518	,0211	,0989	,1057
100	$m_1$	,0006	,0001	,0001	,0034	,0020	,0046
	$m_2$	,0123	,0090	,0086	,0240	,0325	,0655
	$m_3$	,0480	,0447	,0442	,0687	,0771	,1101
	$m_4$	,0272	,0158	,0146	,0600	,0358	,0080
	$m_5$	,0041	,0397	,0521	,0102	,0500	,0977
200	$m_1$	,0012	7e-05	7e-05	,0008	,0002	,0034
	$m_2$	,0114	,0065	,0070	,0125	,0084	,0577
	$m_3$	,0359	,0310	,0315	,0403	,0362	,0855
	$m_4$	,0352	,0108	,0094	,0293	,0186	,0069
	$m_5$	-,0009	,0342	,0506	,0015	,0152	,0878

Table 3: Criteria values for designs 5 and 6

We will apply our method with boundary correcting kernels on the transformed data  $y_i = \hat{T}(x_i)$ , ( $i = 1, \dots, n$ ), where  $\hat{T}(\cdot)$  refers to  $T(\cdot)$  with estimated  $\alpha$  and  $M$ . The same way we define  $\hat{t}(\cdot)$ . The resulting kernel density estimate we call  $\hat{f}_{transf}(y)$ . Then, the final density estimate for the original data is  $\hat{f}(x_i) = \hat{f}_{transf}(y_i) \cdot \hat{t}(x_i)$ . Note that  $\{y_i\}_{i=1}^n \in (0, 1)$ .

So, we have to define a local linear estimator on the interval  $(0, 1)$  As long as all bandwidths considered - both the bandwidth of the original kernel estimator and the bandwidth of the one-sided kernel estimator - are smaller than one half then we can continue to use the local linear estimator defined above that only take care of one boundary. Since this is indeed the case in our two specific applications below, we do not generalize our procedure to be able to take care of two boundaries. Also, as long as all bandwidths are smaller than one half, we can change our above approach to one-sided kernel estimation, by estimating to the right when estimating in the interval  $(0, 1/2)$  and estimating to the left when estimating in the interval  $(1/2, 1)$ . The asymptotic theory that our one-sided kernel bandwidth approach is based on is of course unchanged by



this and we can proceed just as described above. We there replace (3) by

$$\bar{K}(u) = \begin{cases} 2K(u) & \text{if } u < 0 \text{ and } 0 < x \leq 1/2 \\ 2K(u) & \text{if } u > 0 \text{ and } 1/2 < x < 1 \\ 0 & \text{otherwise} \end{cases},$$

calculate our one-sided bandwidth and rescale to our final bandwidth.

## 5.2 Operational Risks: Execution, Delivery and Process Management

We first apply our method to a small data set with sample size  $n = 75$ . It has been taken from a major publicly available financial loss data base on operational risk data. We consider the loss line "Execution, delivery and process management" with relatively few reported claims.

First we have transformed the data along the lines described above. For this transformation  $\hat{T}$  we got from maximum likelihood  $\hat{\alpha} = 1.1129968$  and  $\hat{M} = 3.2357247$ . Then we have searched for the optimal bandwidths on an equispaced grid of 50 bandwidths from  $h_{min} = 1/75$  to  $h_{max} = 0.5$ , respectively  $h_{max} = 0.27 \approx C/2$  for the one sided cross validation, see discussion above. For kernel  $K$  we have used again the Epanechnikov kernel as we did in the simulations. The results were  $\hat{h}_{CV} = 0.05$  for the classical cross validation,  $\hat{h}_{OSCV} = 0.24$  for the one sided cross validation,  $\hat{h}_S = 0.29$  for Silverman's bandwidth, and  $\hat{h}_{PI} = 0.43$  for the refined plug-in method. Silverman's bandwidth has been calculated as described in Section 4  $g_S$  but corrected for Epanechnikov kernels. Compared to the other bandwidth estimates,  $\hat{h}_{PI}$  is much to big. A closer look at the calculation revealed that for this (small) data set the refined plug-in method has serious problems with the estimation of the second derivative  $f''$ .

In Figure 2 are given the resulting density plots for both the original and the transformed data except for  $\hat{h}_{PI}$  as this was ridiculously oversmoothed. It can be seen clearly that CV tends to under- and plug-in to oversmooth whereas our one sided cross validation method lies in between. While the difference between the three curves might seem slight to the untrained eye, the difference in the heaviness of the three tails are actually enormous and economic judgements would be very different for these three curve estimators. Note that the transformation approach allow us to compare the entire tail and that we therefore are able to get a visual impression of the relationship between

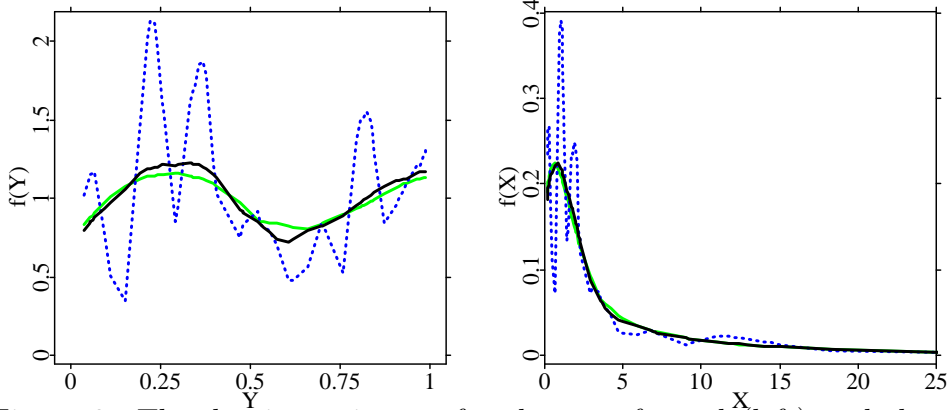


Figure 2: The density estimates for the transformed (left) and the original data (right): black solid for  $\hat{h}_{OSCV}$ , blue dashed for  $\hat{h}_{CV}$ , and green solid for  $\hat{h}_S$ . The graph on the right is cut off at  $x = 25$ .

tails for different estimators. This visual comparison is more complicated on the original scale: we can not capture the entire positive axis in one picture.

In the second example we consider external fraud taken from the same financial data base as our first data set was taken from. Here the size of the data set is  $n = 538$ .

For this data set, the transformation  $\hat{T}$  has been performed with the maximum likelihood estimates  $\hat{\alpha} = 1.113242$  and  $\hat{M} = 4.0$ . Accordingly to the sample size we have searched for the optimal bandwidths on an equispaced grid of 50 bandwidths from  $h_{min} = 1/538$  to  $h_{max} = 0.25 < C/2$ . Here, we also tried with larger  $h_{min}$  for reasons we discuss below. For kernel  $K$  we have used again the Epanechnikov kernel. The results were that  $\hat{h}_{CV}$  (using the classical cross validation) always converged to zero whatever  $h_{min}$  has been. Now it is well known that for increasing  $n$  cross validation suffers from this problem. As a remedy, Feluch and Koronacki (1992) propose to leave out in the cross validation not just observation  $x_i$  ( $y_i$  respectively in our application) but  $x_i$  and its neighbors, in other words an  $\epsilon$ -environment  $U_\epsilon(x_i)$ . However, in practice it is not clear how large this environment should be, this depends certainly on sample size  $n$ . Moreover, often  $\hat{h}_{CV}$  varies a lot with the size of  $U_\epsilon(x_i)$ . Therefore, in this application we have given in the search for a reasonable  $\hat{h}_{CV}$ . Further results have been  $\hat{h}_{OSCV} = 0.100$  for the one sided cross validation,  $\hat{h}_S = 0.190$  for Silverman's bandwidths, and  $\hat{h}_{PI} = 0.214$  for the refined plug-in method.

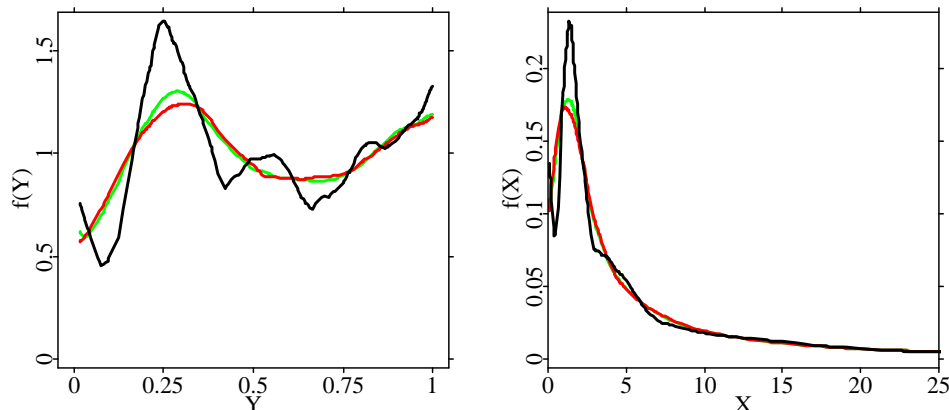


Figure 3: The density estimates for the transformed (left) and the original data (right): black solid for  $\hat{h}_{OSCV}$ , green solid for  $\hat{h}_S$ , and red solid for  $\hat{h}_{PI}$ . The graph on the right is cut off at  $x = 25$ .

In Figure 3 are given the resulting density plots for both the original and the transformed data. Again, obviously CV tends to strongly undersmooth (not plotted as  $\hat{h}_{CV} \approx 0.0$ ) and plug-in to oversmooth whereas our one sided cross validation method lies in between although with the tendency to undersmooth. Also here the difference of the two curves might seem slight - standard cross-validation is left out since it did not work at all in this case - the difference in the heaviness of the two are so big that it is very important which of these curves economic judgements are based on.

We conclude that our one-sided method beats clearly classic cross validation and refined plug-in in both the simulations and the real data examples. We have seen one example in which refined plug-in breaks down, and one in which classic cross validation breaks down whereas our method does reasonably well throughout.

## References

- Bowman, A., 1984, An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, **71**, 353–360.
- Buch-Larsen, T., Nielsen, J.P., Guillén, M., and Bolancé, C., 2005, Kernel density estimation for heavy-tailed distributions using the champernowne transformation. *Statistics*, **39**, 503-518.

- Champernowne, D.G. (1936). The Oxford meeting, September 25-29, 1936, *Econometrica*, Vol. 5, No. 4, October 1937.
- Champernowne, D.G. (1952). The graduation of income distributions. *Econometrica*, **20**, 591-615.
- Cheng, M.Y., 1997a, Boundary-aware estimators of integrated squared density derivatives. *Journal of the Royal Statistical Society Ser. B*, **50**, 191-203.
- Cheng, M.Y., 1997b, A bandwidth selector for local linear density estimators. *The Annals of Statistics*, **25**, 1001-1013.
- Chiu, S.T., 1991, Bandwidth selection for kernel density estimation. *The Annals of Statistics*, **19**, 1883-1905.
- Clements, A.E., Hurn, A.S. and Lindsay, K.A., 2003, Möbius-like mappings and their use in kernel density estimation. *Journal of the American Statistical Association*, **98**, 993-1000.
- Feluch, W., and Koronacki, J., 1992, A note on modified cross-validation in density estimation. *Computational Statistics & Data Analysis*, **13**, 143-151.
- Hall, P., 1983, Large sample optimality of least squares cross-validation in density estimation. *The Annals of Statistics*, **11**, 1156-1174.
- Hall, P. and Marron, J.S., 1987, Extent to which Least-Squares Cross-Validation Minimises Integrated Square Error in Nonparametric Density Estimation. *Probability Theory and Related Fields*, **74**, 567-581.
- Hall, P., Marron, J.S. and Park, B., 1992, Smoothed cross-validation. *Probability Theory and Related Fields*, **92**, 1-20.
- Härdle, W., Hall, P. and Marron, J.S., 1988, How far are automatically chosen regression smoothing parameters from their optimum?. *Journal of the American Statistical Association*, **83**, 86-99.
- Hart, J.D. and Yi, S., 1998, One-Sided Cross-Validation. *Journal of the American Statistical Association*, **93**, 620-631.
- Jones, M.C., 1993, Simple boundary correction in kernel density estimation. *Statistics and Computing*, **3**, 135-146.

- Rudemo, M., 1982, Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, **9**, 65–78.
- Scoot, D.W. and Terrell, G.R., 1987, Biased and Unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, **82**, 1131–1146.
- Sheather, S.J. and Jones, M.C., 1991, A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Ser. B*, **53**, 683–690.
- Wand, M.P. and Jones, M.C., 1995, *Kernel Smoothing*, Vol. 60 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, London.
- Yi, S., 1996, On one-sided cross-validation in nonparametric regression. *Unpublished Ph.D. dissertation*, Texas A&M University, Dept. of Statistics.
- Yi, S., 2001, Asymptotic stability of the OSCV smoothing parameter selection. *Communications in Statistics. Theory and Methods*, **30**, 2033–2044.
- Yi, S., 2004, A comparison of Two Bandwidth Selectors OSCV and AICc in Nonparametric Regression. *Communications in Statistics. Simulation and Computation*, **34**, 585–594.